

JC890 U.S. PTO  
08/04/00

08.07.00

A

DOCKET NO. : MERK-0004/20671

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:

Bert Gunter

Serial No.: Not Yet Assigned

Group Art Unit: Not Yet Assigned

Filing Date: Herewith

Examiner: Not Yet Assigned

For: METHOD AND APPARATUS FOR POSITIONALLY CORRECTING DATA IN  
A THREE DIMENSIONAL ARRAY

EXPRESS MAIL LABEL NO: EL531431154US  
DATE OF DEPOSIT: August 4, 2000

Box ☒ Patent Application  
☐ Provisional ☐ Design

Assistant Commissioner for Patents  
Washington DC 20231

Sir:

PATENT APPLICATION TRANSMITTAL LETTER

Transmitted herewith for filing, please find

☒ A Utility Patent Application under 37 C.F.R. 1.53(b).

It is a continuing application, as follows:

☐ continuation ☐ divisional ☐ continuation-in-part of prior application number  
\_\_\_\_/\_\_\_\_.

☐ A Provisional Patent Application under 37 C.F.R. 1.53(c).

☐ A Design Patent Application (submitted in duplicate).

Including the following:

JC781 U.S. PTO  
09/632422  
08/04/00

☐ Provisional Application Cover Sheet.

☒ New or Revised Specification, including pages 1 to 48 containing:

☒ Specification

☒ Claims

☒ Abstract

☐ Substitute Specification, including Claims and Abstract.

☐ The present application is a continuation application of Application No. \_\_\_\_\_ filed \_\_\_\_\_. The present application includes the Specification of the parent application which has been revised in accordance with the amendments filed in the parent application. Since none of those amendments incorporate new matter into the parent application, the present revised Specification also does not include new matter.

☐ The present application is a continuation application of Application No. \_\_\_\_\_ filed \_\_\_\_\_, which in turn is a continuation-in-part of Application No. \_\_\_\_\_ filed \_\_\_\_\_. The present application includes the Specification of the parent application which has been revised in accordance with the amendments filed in the parent application. Although the amendments in the parent C-I-P application may have incorporated new matter, since those are the only revisions included in the present application, the present application includes no new matter in relation to the parent application.

☐ A copy of earlier application Serial No. \_\_\_\_\_ Filed \_\_\_\_\_, including Specification, Claims and Abstract (pages 1 - @@), to which no new matter has been added TOGETHER WITH a copy of the executed oath or declaration for such earlier application and all drawings and appendices. Such earlier application is hereby incorporated into the present application by reference.

☐ Please enter the following amendment to the Specification under the Cross-Reference to Related Applications section (or create such a section) : "This Application:  
☐ is a continuation of ☐ is a divisional of ☐ claims benefit of U.S. provisional Application Serial No. \_\_\_\_\_ filed \_\_\_\_\_  
\_\_\_\_\_

- 
- 
- ☐ Signed Statement attached deleting inventor(s) named in the prior application.
- ☐ A Preliminary Amendment.
- ☒ Two Sheets of ☒ Formal ☐ Informal Drawings.
- ☐ Petition to Accept Photographic Drawings.
- ☐ Petition Fee
- ☒ An ☐ Executed ☒ Unexecuted Declaration or Oath and Power of Attorney.
- ☐ An Associate Power of Attorney.
- ☐ An ☐ Executed ☐ Copy of Executed Assignment of the Invention to \_\_\_\_\_
- ☐ A Recordation Form Cover Sheet.
- ☐ Recordation Fee - \$40.00.
- ☐ The prior application is assigned of record to \_\_\_\_\_
- ☐ Priority is claimed under 35 U.S.C. § 119 of Patent Application No. \_\_\_\_\_  
filed \_\_\_\_\_ in \_\_\_\_\_ (country).
- ☐ A Certified Copy of each of the above applications for which priority is claimed:
- ☐ is enclosed.
- ☐ has been filed in prior application Serial No. \_\_\_\_\_ filed \_\_\_\_\_.
- ☐ An ☐ Executed or ☐ Copy of Executed Earlier Statement Claiming Small Entity Status under 37 C.F.R. 1.9 and 1.27
- ☐ is enclosed.
- ☐ has been filed in prior application Serial No. \_\_\_\_\_ filed \_\_\_\_\_, said status is still proper and desired in present case.

- ☐ Diskette Containing DNA/Amino Acid Sequence Information.
- ☐ Statement to Support Submission of DNA/Amino Acid Sequence Information.
- ☐ The computer readable form in this application \_\_\_\_\_, is identical with that filed in Application Serial Number \_\_\_\_\_, filed \_\_\_\_\_. In accordance with 37 CFR 1.821(e), please use the ☐ first-filed, ☐ last-filed or ☐ only computer readable form filed in that application as the computer readable form for the instant application. It is understood that the Patent and Trademark Office will make the necessary change in application number and filing date for the computer readable form that will be used for the instant application. A paper copy of the Sequence Listing is ☐ included in the originally-filed specification of the instant application, ☐ included in a separately filed preliminary amendment for incorporation into the specification.
- ☐ Information Disclosure Statement.
- ☐ Attached Form 1449.
- ☐ Copies of each of the references listed on the attached Form PTO-1449 are enclosed herewith.
- ☐ A copy of Petition for Extension of Time as filed in the prior case.
- ☐ Appended Material as follows: \_\_\_\_\_.
- ☒ Return Receipt Postcard (should be specifically itemized).
- ☐ Other as follows: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_.

## FEE CALCULATION:

- ☐ Cancel in this application original claims \_\_\_\_\_ of the prior application before calculating the filing fee. (At least one original independent claim must be retained for filing purposes.)

			SMALL ENTITY		NOT SMALL ENTITY	
			RATE	FEE	RATE	FEE
PROVISIONAL APPLICATION			\$75.00	\$	\$150.00	\$
DESIGN APPLICATION			\$155.00	\$	\$310.00	\$
UTILITY APPLICATIONS BASE FEE			\$345.00	\$	\$690.00	\$690.00
UTILITY APPLICATION; ALL CLAIMS CALCULATED AFTER ENTRY OF ALL AMENDMENTS						
	No. Filed	No. Extra				
TOTAL CLAIMS	58- 20 =	38	\$9 each	\$	\$18 each	\$684.00
INDEP. CLAIMS	7- 3 =	4	\$39 each	\$	\$78 each	\$312.00
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM			\$130	\$	\$260	\$0
ADDITIONAL FILING FEE				\$		\$0
TOTAL FILING FEE DUE				\$		\$1686.00

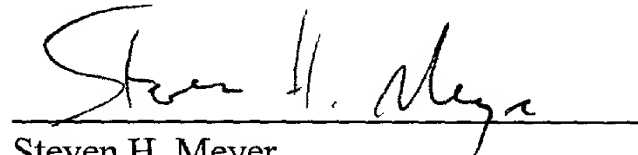
- ☒ A Check is enclosed in the amount of \$ 1686.00.
- ☒ The Commissioner is authorized to charge payment of the following fees and to refund any overpayment associated with this communication or during the pendency of this application to deposit account 23-3050. This sheet is provided in duplicate.
- ☐ The foregoing amount due.
- ☒ Any additional filing fees required, including fees for the presentation of extra claims under 37 C.F.R. 1.16.
- ☒ Any additional patent application processing fees under 37 C.F.R. 1.17 or 1.20(d).
- ☐ The issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance.
- ☒ The Commissioner is hereby requested to grant an extension of time for the appropriate length of time, should one be necessary, in connection with this filing or any future filing submitted to the U.S. Patent and Trademark Office in the above-

identified application during the pendency of this application. The Commissioner is further authorized to charge any fees related to any such extension of time to deposit account 23-3050. This sheet is provided in duplicate.

**SHOULD ANY DEFICIENCIES APPEAR** with respect to this application, including deficiencies in payment of fees, missing parts of the application or otherwise, the United States Patent and Trademark Office is respectfully requested to promptly notify the undersigned.

Date:

August 4, 2000



Steven H. Meyer

Registration No. 37,189

Woodcock Washburn Kurtz  
Mackiewicz & Norris LLP  
One Liberty Place - 46th Floor  
Philadelphia PA 19103  
Telephone: (215) 568-3100  
Facsimile: (215) 568-3439

EL531431154 US

5

10

**Title of the Invention**

Method and Apparatus for Positionally Correcting Data  
in a Three Dimensional Array

15

**Cross-Reference to Related Application**

This application claims the benefit of U.S. Patent Application No. 60/217,772, filed July 12, 2000 and entitled "METHOD AND APPARATUS FOR POSITIONALLY CORRECTING DATA IN A THREE DIMENSIONAL ARRAY", hereby incorporated by reference.

20

**Field of the Invention**

The present invention relates generally to an apparatus for and method of positionally correcting data in a three dimensional array in data sets that are generated from output functions which are subject to variations due to positional effects.

25

**Background of the Invention**

Drug discovery is often achieved through *in vitro* assays employed to identify compounds that have effects on various specific biological processes.

Efforts have been undertaken to identify agents that may block, reduce, or even enhance the interactions between biological molecules.

It is well known that the interaction between a receptor and its ligand often may result, either directly or through some downstream event, in either a deleterious or beneficial effect on a biological system which can thus affect a patient with a condition, disease or disorder associated with activity in such biological system. Accordingly, agents which can reduce, block or enhance interaction between a receptor and its ligand are sought as pharmacologically active entities.

Similarly, it is well known that the enhancement or inhibition of enzyme activities often may result, either directly or through some downstream event, in either a deleterious or beneficial effect on a clinically relevant biological system. Accordingly, efforts are undertaken to identify compounds that serve as substrates, inhibitors or catalysts for enzymatic reactions using *in vitro* assays.

In addition to these examples, there are many other drug targets and biological systems for which *in vitro* assays can and are used to identify pharmacologically active and biologically active agents. For example, human genome research has also uncovered large numbers of new target molecules against which the efficacy of test compounds may be screened.

One strategy employed in modern drug discovery is to maximize the throughput of the assays that are used to screen test agents which possess a desired pharmacological activity. In particular, by screening a large number of different test agents, the probability of testing and identifying a compound with the desired activity is greater. Using robotics and other automation technology together with automated detection systems, high throughput screening assays have been developed which apply industrial / manufacturing concepts and design to research protocols.

High throughput screening assays provide for the performance of multiple identical assays in parallel on a platform matrix. Multiple parallel assays are



performed sequentially. By sequentially performing multiple parallel assays, the data outputted may be compiled and arranged in a three dimensional array.

Sequential and parallel processing in high throughput screening methods using robotics and automation allows for the ability to test many thousands of test agents in an assay. Automated screening procedures allow high throughput evaluation of individual test agents in collections or libraries which contain large numbers of test agents in order to assess functional biological / pharmacological properties of each test agent.

Screening of collections of test agents is an important aspect of efforts to identify lead compounds that have pharmacological activity and that can be further developed into new drugs and therapeutic compositions. Such test agents include but are not limited to chemically synthesized molecules, including libraries of compounds synthesized by combinatorial chemistry; natural products, including cells, cell extracts, nucleic acid molecules, cell culture media, proteins, isolated genetic material, fungal extracts and microbial fermentation broths; and recombinant products such as viral and phage particles, proteins and peptide libraries. Generally, the type of test agent used in the high throughput screening includes any composition or molecule which can be used in an *in vitro* assay. Most commonly, high throughput screening is performed using collections, also referred to as libraries, of test agents that include thousands of individual chemical entities or compositions.

High throughput screening procedures provide the ability to perform large numbers of identical functional assays that are predictive of bioactivity in a fully integrated automated format that accelerates data collection and lead identification, while also cutting costs. Each hit, i.e. test agent that produces a positive assay result, represents a lead candidate compound that has pharmacological activity. Lead candidate compounds may then be further investigated for development.

The assays used in high throughput screening procedures include any detectable activity with pharmacological or biological significance. Activities

which are assessed in high throughput screening procedures include, but are not limited to, enzyme activation, enzyme inhibition, ligand-receptor binding, ligand-receptor binding inhibition, cell cycle inhibition, cell cycle activation, cell growth, cell division, cell activation, cell inhibition, activation of production of and/or  
5 release of cellular factors, inhibition of production of and/or release of cellular factors, ion pump, transport or channel activity, ion pump, transport or channel inhibition, activation of DNA synthesis, inhibition of DNA synthesis, activation of RNA synthesis, inhibition of RNA synthesis, activation of protein synthesis, inhibition of protein synthesis, metabolic activity, metabolic inhibition, activation of  
10 apoptosis, and inhibition of apoptosis.

High throughput screening procedures can be used to identify agents useful to treat infectious diseases such as compounds with anti-viral activity such as those that inhibit viral attachment to cells, infection, viral gene expression, viral gene replication, and viral particle assembly; antibiotic activity  
15 including anti-bacterial activity, anti-fungal activity, and anti-parasitic pathogen activity. High throughput screening assays are used in the search to identify pharmacologically active agents for use in medical and nutritional treatments and regimens such as, but not limited to, anti-cancer agents, anti-inflammatory agents, immunosuppressive agents, neuropharmacologically active agents, blood  
20 chemistry modifying agents, and agents for treatment of cardiac, pulmonary, renal, hepatic, pancreatic, bone, blood, gastrointestinal, and dermatological diseases.

Those skilled in the art routinely apply high throughput screening technology to identify active agents in a variety of different chemical and biological  
25 systems using a variety of target and reactions. Although drug discovery is a common application of high throughput screening, the present invention is useful in the field of high throughput screening data analysis generally.

In such assays, activation and inhibition can be detected and measured by detection and/or measurement of various detectable markers, such  
30 as but not limited to, those which are detected by their radioactivity,

characteristics which can be observed optically or by electromagnetic detection, scintillation counting, fluorescence, visible dye changes in intracellular concentration of ionized calcium, cAMP or pH, trans-membrane potential and other physiological and biochemical characteristics of living cells which can be  
5 measured by a variety of conventional means, for example using specific fluorescent, luminescent or color developing dyes and the like.

High throughput screening is often performed using collections of test agents that are individually dispensed in wells of multi-well plates. Standard plates usually contain 96 wells (organized into an 8 X 12 array) while some larger  
10 plate sizes contain 384 wells (16 X 24 array), 1536 wells (32 X 48 array) and 3456 wells (48 X 72 array). Although these configurations are among the most common employed, other arrangements are equally useful in the present invention. Likewise, microchip array technology provides for the deposition of libraries of combinatorial chemical and biological materials in fixed two-  
15 dimensional arrays. Importantly, whether the platform is a microtitre plate, a microchip array or some other platform for performing parallel assays on a collection of individual test agents, the assay sites which contain test agents are arranged in identical matrices.

To assess the activity or inhibitory effect that a test agent has in an  
20 assay, positive and negative control samples are provided. Such controls are test samples that include the presence or absence of an active compound. These positive and negative controls provide data to which the results of test assays can be compared in order to determine the activity of the test agents.

As may be appreciated, several problems are associated with the  
25 analysis of data generated in high throughput screening assays. Such problems arise from variability in controls, variability in samples, and systemic variability, among other things.

A significant problem experienced in high throughput screening is that of positional effects which exist with respect to the location of wells on each  
30 plates. That is, a variability of background exists that may be associated with

specific well locations within the matrix of a plate, where such variability is substantially consistent across a series of plates.

The variability of background due to positional effects has been observed to be sufficient to be responsible for a significant number of false positives and false negatives relative to controls. That is, data from test assays at specific well locations is consistently identified as being higher or lower for the parameter measured relative to positive and negative control data when compared to corresponding data from other well locations. The phenomenon of positional effects based upon well location on plates represents a real and significant problem to the predictability of data from high throughput screening.

False negatives result in the failure to identify a lead candidate compound which has the desired pharmacological activity from the library of chemical compounds being tested. Such failure to identify is of course a missed opportunity to further examine the compound and perhaps identify pharmacological relevance for the compound.

False positives result in further investigation and development of a compound which does not actually have the desired pharmacological activity. The further testing of false positives is an ineffective use of manpower and other resources as well as a waste of valuable stock from a chemical collection.

Accordingly, there is a need for an improved method and apparatus for analyzing high throughput screening data, a need for an improved method and apparatus for reducing the number of false positives and false negatives in high throughput screening assays, a need for an improved method and apparatus for correcting high throughput screening data for positional effects, and a need for an improved method and apparatus for analyzing assay conditions in high throughput screening using data analysis.

### Summary of the Invention

The present invention satisfies the aforementioned need by providing a method of obtaining and evaluating assay data. In the method, an

assay is performed such that a compendium of raw assay data is developed. The raw assay data is compensated for systematic and positional effects, the compensated data is scored, and the scored data is formatted according to a determined format.

5                   In addition, the present invention provides a method of positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates  $p$ . Each plate  $p$  has a plurality of wells organized into rows  $i$  and columns  $j$ . Each well  $(i, j, p)$  has a raw value  $x_{ijp}$  associated therewith, where the raw values  $x_{ijp}$  comprise the raw assay data. Each raw value  $x_{ijp}$  of an associated  
10 well  $(i, j, p)$  is deconstructed into:

a plate effect value representing extraneous effects  
attributable to the plate  $p$  of the well  $(i, j, p)$ ;

a row effect value representing extraneous effects  
attributable to the row  $i$  on the plate  $p$  of the well  $(i, j, p)$ ;

15                   a column effect value representing extraneous effects  
attributable to the column  $j$  on the plate  $p$  of the well  $(i, j, p)$ ;

a non-additive, interaction effect representing an extraneous  
positional effect beyond the plate, row, and column effects previously determined  
for the  $(i, j, p)$  well on plate  $p$ ; and

20                   a residual data value that is left over once all the above  
extraneous effects are taken into account.

Thereafter, the residual data value associated with each well  $(i, j, p)$  is employed to represent the well  $(i, j, p)$  as compared with all other wells  $(i, j, p)$  on the plate  $p$ .

## 25                   Brief Description of the Drawings

The foregoing summary as well as the following detailed description of the present invention will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there are shown in the drawings embodiments which are presently preferred. As should be

understood, however, the invention is not limited to the precise arrangements and instrumentalities shown. In the drawings:

Fig. 1 is a flow chart detailing steps performed in formulating assay results in accordance with one embodiment of the present invention;

5 Fig. 2 is a block diagram showing a computer on which the steps detailed in Fig. 3 may be performed; and

Fig. 3 is a flow chart detailing steps performed in positionally correcting assay data in accordance with one embodiment of the present invention.

10

**Detailed Description of Preferred Embodiments**

Certain terminology may be used in the following description for convenience only and is not considered to be limiting. For example, the words “left”, “right”, “upper”, and “lower” designate directions in the drawings to which reference is made. Likewise, the words “inwardly” and “outwardly” are directions toward and away from, respectively, the geometric center of the referenced object. The terminology includes the words above specifically mentioned, derivatives thereof, and words of similar import.

15

Generally, in the present invention, assay screening data is positionally corrected and then formatted into a form where such positionally corrected data may be presented to an appropriate assay analyst or the like. Thus, and referring to Fig. 1, now, the assay is performed such that a compendium of raw assay data is developed (step 101). Such raw assay data may be in any appropriate form without departing from the spirit and scope of the present invention. For example, such raw assay data may be expressed in terms of its original units of measure, or may be scaled based on an appropriate numerical scale (0 - 1, 0 - 100, -100 - +100, etc.).

20

25

Preferably, the raw data is in a computer-readable form for purposes of allowing a computer-based algorithm to operate on such data, although such raw data may also be transcribed or otherwise converted into a computer-

30

readable form without departing from the spirit and scope of the present invention.

Any appropriate computer-readable form may be employed without departing from the spirit and scope of the present invention. For example, the computer-readable form may be an ASCII delimited file, a spreadsheet file, a table file, a database file, or the like. Presumably, the computer-readable form of the raw assay data is accessible by any particular software employed to perform the algorithm, as described below.

Once the raw assay data is developed and is in the computer-readable form, an appropriate algorithm is employed to process the raw assay data and compensate such raw assay data for systematic and/or positional effects (step 103). Such algorithm is described in more detail below in connection with Fig. 3. In the processing and compensating, the algorithm estimates background effects such as those that derive from a well being on a particular plate, being in a particular row on a plate, being in a particular column on a plate, being in a particular part of a series of plates, etc. In addition, the algorithm adjusts the compensated raw data for variations from plate to plate to result in a score value for each well. Note that although the algorithm as disclosed herein orients wells in terms of rows and columns on a plate, any other appropriate orientation system may be employed without departing from the spirit and scope of the present invention.

Finally, now that all systematic / positional / background effects have been removed from the raw assay data and such raw assay data has been scored to result in a score value for each well, such score values for all the wells may then be compared / organized / ranked and otherwise formatted into an appropriate form (step 105). Any appropriate formatting form may be employed without departing from the spirit and scope of the present invention. For example, each well may be ranked in a list according to its potency as represented by the score value for such well. Such formatted score values are then available for presentation to an appropriate assay analyst or the like.

Positionally Correcting Algorithm:

In one embodiment of the present invention, for purposes of performing the algorithm by which positional correcting takes place, the assay is performed in connection with a series of plates, where each plate is serially  
5 assayed. Thus, each plate has a time aspect or is 'longitudinally' positioned with respect to the other plates. Accordingly, each plate  $p$  in the series of plates is indexed by its order within the series:

$$p = 1, 2, \dots, P.$$

10

Likewise, for each plate  $p$ , each row  $i$  thereon and each column  $j$  thereon is indexed by its respective order on the plate:

$$i = 1, 2, \dots, I; \text{ and}$$

15

$$j = 1, 2, \dots, J.$$

20

Thus, the raw measured data value as obtained from the assay for any particular well at row  $i$  and column  $j$  on plate  $p$  is  $x_{ijp}$ . Of course, any other appropriate positional identification system may be employed without departing from the spirit and scope of the present invention.

25

In one embodiment of the present invention, the raw measured data values from the assay are employed to fit each raw measured data value to a model. In the model, each raw measured data value  $x_{ijp}$  from a well  $(i, j, p)$  is deconstructed into:

30

- a plate effect value representing extraneous effects attributable to the plate  $p$  of the well  $(i, j, p)$ ;
- a row effect value representing extraneous effects attributable to the row  $i$  on the plate  $p$  of the well  $(i, j, p)$ ;
- a column effect value representing extraneous effects attributable



to the column j on the plate p of the well (i, j, p);

- a non-additive, interaction effect representing extraneous positional effects attributable to consistent positional effects beyond the plate, row, and column effects previously determined for the (i, j, p) well on plate p; and

- a residual data value that is left over once all the above extraneous effects are taken into account.

As should be appreciated, the residual data value more truly represents the potency of the sample in the well (i, j, p) as compared with all other wells (i, j, p) on the plate p. Accordingly, the purpose of the model is to obtain such residual data value.

Re-stated in more mathematical terms, the model fit by the algorithm of the present invention is:

$$x_{ijp} = \mu_p + R_{ip} + C_{jp} + \text{smooth}_p(e_{ijp}) + \varepsilon_{ijp}$$

where  $x_{ijp}$  is the aforementioned raw measured data value for the well at row i and column j on plate p, and where  $\varepsilon_{ijp}$  is the possible systematic interaction effect for the (i, j) well on plate p. That is,  $\varepsilon_{ijp}$  is 'the residual data' left over after all possible positional effects (discussed immediately below) have been removed from the raw measured data value. Note that all the non-potent  $\varepsilon_{ijp}$  values for a plate p – expected to be the bulk of the data – are expected to vary about 0 in a generally Gaussian manner, while the potent  $\varepsilon_{ijp}$  values will differ greatly from 0.

The element  $\mu_p$  in the above equation represents the overall median of all the raw measured data values of the wells on plate p (i.e. 'the plate median'). Thus, the plate median represents the possible systematic measurement plate offset for plate p. Similarly,  $R_{ip}$  is the median of all the raw values of the wells for row i on plate p (i.e., 'the row effect') after taking into consideration the plate median, and  $C_{jp}$  is the median of all the raw values of the wells for column j on plate p (i.e., 'the column effect') after taking into

consideration the plate median. Thus, the row and column effects represent the possible systematic measurement row offset for row  $i$  on plate  $p$  and the possible systematic measurement column offset for column  $j$  on plate  $p$ .

The element  $\text{smooth}_p(e_{ijp})$  is the possible systematic measurement longitudinal offset for the  $(i, j)$  well on plate  $p$  (i.e., 'the longitudinal effect' and 'the non-additive interaction'). As will be discussed in more detail below,  $\text{smooth}_p(e_{ijp})$  results from a smoothing function, and is employed to take into account longitudinal position effects by combining data over similar plates to determine this effect. That is, it is expected that systematic positional effects like edge effects or a 'high' region on a plate would be fairly consistent from plate to plate, especially for plates that are measured close together in time sequence. Therefore, the model takes advantage of the information in "nearby" plates to "average" results from wells in the same  $(i, j)$  position after correcting for the additive effects of plate, row and column so that the corrected data are expected to be effectively similar up to measurement error.

The underlying assumption incumbent in the above-specified model is that almost all the wells  $(i, j, p)$  can be assumed to contain zero or low potency compounds, with only a small proportion of wells  $(i, j, p)$  containing high potency compounds of interest. Accordingly, resistant statistical methods that ignore the high potency 'outliers' are used to fit the model. The fitted model thus should capture the systematic positional measurement effects incumbent in any assay, while the 'residual data'—the  $\varepsilon_{ijp}$ 's—should contain the leftover non-systematic noise, including the aforementioned high potency outliers.

Commercially available statistical software may be employed to implement many of the functions of the algorithm of the present invention. Such software may for example include S-PLUS statistical data analysis software, produced and/or marketed by MATHSOFT, Inc. of Cambridge, Massachusetts, although any other appropriate software may be employed without departing from the spirit and scope of the present invention. Examples of S-PLUS code written for the S-PLUS software and employed to implement the positionally correcting

algorithm of the present invention are set forth in the attached Appendix. As should be appreciated by the relevant public, such S-PLUS software and other similar software include analyzing procedures as discussed for example in Exploratory Data Analysis, Tukey, John W., Addison-Wesley: Reading, MA.

5 (1977), which is hereby incorporated by reference.

Referring now to Fig. 2, such software may be operated in the form of modules or otherwise on any appropriate computer 10 without departing from the spirit and scope of the present invention. As is typical, such computer 10 may include appropriate computer components including a data entry device such as a  
10 modem or network connection 12, a keyboard 14, a data viewing device such as a screen 16, a processor 18, and memory 20, among other things.

In any case, the algorithm of the present invention proceeds as follows. Preliminarily, the raw measured data values  $x_{ijp}$  from the sample wells of all the plates are inputted / received into a data structure 22 in the memory 20 of  
15 the computer 10 (step 301 of Fig. 3). Such data structure 22 may be any appropriate data structure without departing from the spirit and scope of the present invention.

Thereafter, for each plate  $p$ , the raw data  $x_{ijp}$  for such plate  $p$  is resistantly fit to a row-column additive model (step 303):

20

$$y_{ijp} = \mu_p + R'_{ip} + C'_{jp} + e_{ijp} .$$

where:

- 25  $y_{ijp}$  = the raw measured data value for the well at row  $i$  and column  $j$  on plate  $p$ , as obtained from the data structure of step 301;
- $\mu_p$  = the overall "average" for plate  $p$  (i.e. 'plate effect'), as computed;
- $R'_{ip}$  = the possible systematic measurement row offset for row  $i$  on  
30 plate  $p$  (i.e., 'row effect'), as computed;

$C'_{jp}$  = the possible systematic measurement column offset for column  $j$  on plate  $p$  (i.e., 'column effect'), as computed; and

$e_{ijp}$  = the residual data without taking into account any longitudinal / interactive effect.

5

In one embodiment of the present invention, the Tukey two way resistant median polish procedure is employed for such resistant fit, although any other appropriate procedure may be employed without departing from the spirit and scope of the present invention. As is known, such Tukey median polishing procedure is coded into and available from the aforementioned S-PLUS software. Accordingly, since such procedure is known to those in the relevant public, further discussion and explanation thereof need not be provided herein. Suffice it to say that given the raw measured data values from the data structure, such procedure employs an iterative procedure to result in a standard resistant row / column additive fit,

10 thereby solving for each  $\mu_p$ ,  $R'_{ip}$ ,  $C'_{jp}$ , and  $e_{ijp}$ .

An issue arises, though, in the situation where, for example, there happens to be three potent wells in a column. This is a rare occurrence, but can and does nevertheless happen. Suppose also that such column with three potent wells has eight wells total, two of which are empty. Thus, such column has six

20 wells containing samples, three of which have outliers representing potent samples. In such a situation, the  $C'_{jp}$  column effect for such column would be affected by the outliers, which is not desired. As should now be appreciated, the  $C'_{jp}$  column effect should contain errant column-based positional data, not actual non-positional data representing potent compounds.

25 In one embodiment of the present invention, to assure that multiple outliers in a column or a row do not overly affect column and row effect calculations, each  $R'_{ip}$  and each  $C'_{jp}$  is longitudinally (plate-wise) non-linearly smoothed so that their values in any plate  $p$  cannot be much different from nearby plates (step 305). Assuming that the same hit and miss pattern of values does

30 not repeat in nearby plates, which is essentially a certainty, such longitudinal  $C'_{ip}$

and  $R'_{jp}$  smoothing results in the transfer of potent well effects to the residual data, where they belong.

Non-linear smoothing is known to those in the relevant public, and accordingly further discussion and explanation thereof need not be provided herein. Suffice it to say that given, for example, a series of  $R'_{jp}$ 's from adjacent plates  $P_X, P_{X+1}, P_{X+2}, P_{X+3}, P_{X+4}, P_{X+5}$ , etc.:

$$R'_{jp}, R'_{jp+1}, R'_{jp+2}, R'_{jp+3}, R'_{jp+4}, R'_{jp+5}, \text{ etc.},$$

10 the smoothed  $R'_{jp}$  ( $R_{jp}$ ) may for example be the median of  $R'_{jp-1}, R'_{jp}$ , and  $R'_{jp+1}$ . As may be appreciated, other smoothing functions, both simple and complex, may be employed. In fact, any appropriate smoothing function may be employed without departing from the spirit and scope of the present invention.

The amount of longitudinal smoothing of row and column effects  
15 necessary depends on the unknown true situation, and is therefore difficult to determine with certainty. However, for present purposes, a minimal amount will suffice because the residuals are themselves longitudinally smoothed, as will be discussed below, and a fairly rough estimate of the row and column effects is all that is needed.

20 In one embodiment of the present invention, a Tukey-type running median smoother is employed for such longitudinal row effect and column effect smoothing, although any other appropriate smoother may be employed without departing from the spirit and scope of the present invention. In one embodiment of the present invention, the smoother is 4(3RSR)2 H with the "twicing" option set  
25 to False. This results in a somewhat "rough" smooth. As is known, such Tukey-type running median smoother is coded into and available from the aforementioned S-PLUS software.

The result of such longitudinal row effect and column effect smoothing is that the un-smoothed  $R'_{ip}$  and  $C'_{jp}$  values in the previous equation  
30 are substituted with smoothed  $R_{ip}$  and  $C_{jp}$  values as calculated by the smoother.

Thus, the residual value  $e_{ijp}$  in the same equation is adjusted by the smoothing to  $e'_{ijp}$ :

$$y_{ijp} = \mu_p + R_{ip} + C_{jp} + e'_{ijp} .$$

5

Once  $e'_{ijp}$  has been derived for each well (i, j) of each plate p, such  $e'_{ijp}$ 's are then non-linearly smoothed across the plates p by plate position. That is, the smoothing process is performed longitudinally for each well (i, j) to approximate any interactive effect (step 307), resulting in:

10

$$y_{ijp} = \mu_p + R_{ip} + C_{jp} + \text{smooth}_p(e'_{ijp}) + r_{ijp}.$$

where  $\text{smooth}_p(e'_{ijp})$  is the possible systematic measurement longitudinal offset (non-additive interactive offset) for the (i, j) well on plate p (i.e., the interactive effect'), and  $r_{ijp}$  is the residual data left over after taking into account any interactive effect, as calculated.

Such interactive effect is approximated by longitudinal smoothing because no replicates are available in that each sample is tested only once in one well. Such approximation is thus achieved by assuming that nearby plates (longitudinally) are "pseudo-replicates" after correcting for their plate, row and column effects and combining the results by longitudinal smoothing. Nevertheless, such longitudinal smoothing is conceptually calculating a background positional effect on the plate p beyond that attributable to the row / column additive effects. Although perhaps 'a cheat', the smoothing works reasonably well to detect and compensate for consistent (over many plates) background positional effects.

Once again, non-linear smoothing is known to those in the relevant public, and accordingly further discussion and explanation thereof need not be provided herein. Suffice it to say that given, for example, a series of  $e'_{ijp}$ 's from adjacent plates  $P_x, P_{x+1}, P_{x+2}, P_{x+3}, P_{x+4}, P_{x+5}$ , etc.:

30

$$e'_{ijp}, e'_{ijp+1}, e'_{ijp+2}, e'_{ijp+3}, e'_{ijp+4}, e'_{ijp+5}, \text{etc.},$$

the smoothed  $e'_{ijp}$  ( $r_{ijp}$ ) may for example be the median of  $e'_{ijp-1}$ ,  $e'_{ijp}$ , and  $e'_{ijp+1}$ .

- 5 As may be appreciated, other smoothing functions, both simple and complex, may be employed. In fact, any appropriate smoothing function may be employed without departing from the spirit and scope of the present invention.

As may be appreciated, then, the result of step 307 is that each  $e'_{ijp}$  is deconstructed into  $\text{smooth}_p(e'_{ijp})$  and  $r_{ijp}$ . In one embodiment of the present invention, the aforementioned Tukey-type running median smoother is employed for such longitudinal smoothing, although any other appropriate smoother may be employed without departing from the spirit and scope of the present invention. In one embodiment of the present invention, the smoother is 4(3RSR)2 H. As may be appreciated, the advantage of such smoother is that it does not tend to hide outliers – potent wells. In contrast, other commonly used time series filtering techniques that are essentially weighted averaging procedures can and do hide such outliers.

Considering the last equation, now, it is seen that the terms  $\mu_p$ ,  $R_{ip}$ ,  $C_{jp}$ , and  $\text{smooth}_p(e'_{ijp})$  represent the fit and contain the systematic positional effects (plate, row, column, and longitudinal / interactive), if any. Thus,  $r_{ijp}$  is the residual and represents the true relative potency of the well (i, j, p) as compared to all other wells (i, j, p) on the plate p – including extreme potencies of active compounds – without the distortion of the positional effects.

However, to compare potencies across plates p, i.e. across the entire assay, it is necessary to normalize each  $r_{ijp}$ . That is, it must be remembered that all the  $r_{ijp}$  values for a plate p are expected to vary about 0 in a generally Gaussian manner. It must also be remembered, though, that as between plates p, the Gaussian spread can and does differ, and must be taken into account when comparing potencies across such plates. Accordingly, in one embodiment of the present invention, each  $r_{ijp}$  on a plate p is normalized by a

standard deviation value derived from all the  $r_{ijp}$ 's on the plate  $p$  (step 309), thus resulting in a score for the well  $(i, j, p)$  that can be compared across plates  $p$ :

$$\text{score}_{ijp} = r_{ijp} / (\text{standard deviation value})_p .$$

5

The standard deviation value may be any appropriate standard deviation value without departing from the spirit and scope of the present invention. For example, the standard deviation value may be a median absolute deviation from median value multiplied by an appropriate constant to produce an unbiased estimate for a Gaussian distribution. Such median absolute deviation from median value and such multiplication constant are known to those in the relevant art and need not be described herein in any detail.

10

15

Once the  $\text{score}_{ijp}$  has been developed for each well  $(i, j, p)$ , all the  $\text{score}_{ijp}$  values for all the wells  $(i, j, p)$  may then be compared / organized / ranked and otherwise formatted into an appropriate form, as was described above in connection with step 105 of Fig. 1.

20

Any particular method may be employed to choose 'hits' based on the positionally-corrected scores without departing from the spirit and scope of the present invention. One can, of course, arbitrarily set a cutoff, although it is to be noted that such cutoffs vary and frequently change for a variety of reasons even after they have been set for a given assay. In point of fact, there can be no cutoff that is always successful in distinguishing true hits from false positives. That is, there is always some probability of false positives or false negatives based on any chosen cutoff.

25

Whatever hit determining device is used in connection with the positionally corrected scores of the present invention, the point is that the better a score is, the more likely it is that the corresponding sample is a true hit. That is, positionally correcting scores in accordance with the present invention provides a better scoring system to increase the probability of finding true hits as one goes down the list from best to worse. Accordingly, it is recommended in connection

30



with the present invention that the best k% of the positionally corrected scores be confirmed. Typically, k is about 1, although k ideally should be greater assuming it can be afforded, remembering that more elaborate assays are both more expensive and time consuming. Importantly, there is no 'magic cutoff', only statistically unusual results.

In the foregoing description, it can be seen that the present invention comprises a new and useful statistical algorithm for positionally correcting assay screening data. The algorithm corrects for possible plate positional effects, including transitory positional effects due to quality control problems like clogged tips, reader anomalies, and so forth. The algorithm does not require the use of any blank and/or control values and works in the presence of missing values. The algorithm also standardizes corrected raw values across plates, thus allowing values to be ranked across plates. It should be appreciated that changes could be made to the embodiments described above without departing from the inventive concepts thereof. For example, instead of using a running median smoother, a 'lowess' procedure may be employed, as should be appreciated by the relevant public. It should be understood, therefore, that this invention is not limited to the particular embodiments disclosed, but it is intended to cover modifications within the spirit and scope of the present invention as defined by the appended claims.

CLAIMS

5

1. A method of evaluating assay data that is arranged in a three dimensional array, the assay being subject to systematic and positional effects, the method comprising:

10 compensating the raw assay data for the systematic and positional effects;  
scoring the compensated data; and  
formatting the scored data according to a determined format.

15 2. The method of claim 1 wherein an assay is performed to generate a compendium of raw assay data that is then compensated for systematic and positional effects.

20 3. The method of claim 1 wherein the raw assay data is generated from a high throughput screening assay to identify a biologically active agent in a collection of test agents, wherein a biologically active agent is identified by identifying a test agent that generates a data point which statistically deviates from other data points in the formatted scored data.

25 4. The method of claim 1 wherein the raw assay data is generated from a high throughput screening assay to identify a pharmacologically active agent in a collection of test agents,  
wherein  
the high throughput screening assay is selected from the group consisting of enzyme activation, enzyme inhibition, ligand-receptor binding, ligand-receptor binding inhibition, cell cycle inhibition, cell cycle  
30

activation, cell growth, cell division, cell activation, cell inhibition, activation of production of and/or release of cellular factors, inhibition of production of and/or release of cellular factors, ion pump, transport or channel activity, ion pump, transport or channel inhibition, activation of DNA synthesis, inhibition of DNA synthesis, activation of RNA synthesis, inhibition of RNA synthesis, activation of protein synthesis, inhibition of protein synthesis, metabolic activity, metabolic inhibition, activation of apoptosis, and inhibition of apoptosis,

the collection of test agents is selected from the group consisting of: chemically synthesized molecules, natural products, cell extracts, nucleic acid molecules, cell culture media, proteins, isolated genetic material, fungal extracts and microbial fermentation broths; and recombinant products, viral particles, phage particles, proteins and peptide libraries; and

the pharmacologically active agent is identified by identifying a test agent that generates a data point which statistically deviates from other data points in the formatted scored data.

5. The method of claim 1 comprising performing an assay such that a compendium of raw assay data is developed in a computer-readable form for purposes of allowing a computer-based algorithm to operate on such data.

6. The method of claim 1 comprising compensating the raw assay data for systematic and/or positional effects by way of a computer-based algorithm.

7. The method of claim 1 wherein the raw data has row-based aspects, column-based aspects, and non-additive interaction-based aspects, the method comprising compensating the raw assay data for row-based positional effects.

8. The method of claim 1 wherein the raw data has row-based aspects, column-based aspects, and non-additive interaction-based aspects, the method comprising compensating the raw assay data for column-based positional effects.

9. The method of claim 1 wherein the raw data has row-based aspects, column-based aspects, and non-additive interaction-based aspects, the method comprising compensating the raw assay data for longitudinal-based positional effects.

10. A method of positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates  $p$ , each plate  $p$  having a plurality of wells organized into rows  $i$  and columns  $j$ , each well  $(i, j, p)$  having a raw value  $x_{ijp}$  associated therewith, the raw values  $x_{ijp}$  comprising the raw assay data, the method comprising deconstructing each raw value  $x_{ijp}$  of an associated well  $(i, j, p)$  into:

a plate effect value representing extraneous effects attributable to the plate  $p$  of the well  $(i, j, p)$ ;

a row effect value representing extraneous effects attributable to the row  $i$  on the plate  $p$  of the well  $(i, j, p)$ ;

a column effect value representing extraneous effects attributable to the column  $j$  on the plate  $p$  of the well  $(i, j, p)$ ;

a non-additive, interaction effect value representing extraneous positional effects attributable to consistent positional effects beyond the plate, row, and column effects previously determined for the  $(i, j, p)$  well on plate  $p$ ; and

a residual data value that is left over once all the above extraneous effects are taken into account,

the method further comprising employing the residual data value associated with

each well (i, j, p) to represent the well (i, j, p) as compared with all other wells (i, j, p) on the plate p.

11. The method of claim 10 wherein the raw assay data is  
5 generated from a high throughput screening assay to identify a biologically active agent in a collection of test agents, wherein the assay is subject to positional and systemic effects, the raw assay data is arranged in a three dimensional array, a biologically active agent is identified by identifying a test agent that generates a data point which statistically deviates from other data points in the formatted  
10 scored data.

12. The method of claim 10 comprising employing the residual data value associated with each well (i, j, p) to represent the well (i, j, p) as compared with all other wells (i, j, p) on all of the plates p.  
15

13. The method of claim 10 comprising deconstructing each raw value  $x_{ijp}$  of an associated well (i, j, p) into a plate effect value representing the overall median of all the raw values of the wells on plate p.

20 14. The method of claim 13 comprising deconstructing each raw value  $x_{ijp}$  of an associated well (i, j, p) into a row effect value representing the median of all the raw values of the wells for row i on plate p after taking into consideration the plate effect value.

25 15. The method of claim 13 comprising deconstructing each raw value  $x_{ijp}$  of an associated well (i, j, p) into a column effect value representing the median of all the raw values of the wells for column j on plate p after taking into consideration the plate effect value.

16. The method of claim 10 comprising deconstructing each raw value  $x_{ijp}$  of an associated well (i, j, p) into a non-additive, interaction effect value representing an additional possible systematic measurement effect beyond the plate, row, and column effect values previously determined for the (i, j, p) well on  
5 plate p.

17. A method of positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates p, each plate p having a plurality of wells organized into rows i and columns j, each well (i, j, p)  
10 having a raw value  $x_{ijp}$  associated therewith, the raw values  $x_{ijp}$  comprising the raw assay data, the method comprising:

resistantly fitting the raw value  $x_{ijp}$  for each well (i, j) for each plate p to a row-column additive model:

15 
$$y_{ijp} = \mu_p + R'_{ip} + C'_{jp} + e_{ijp} .$$

where:

- 20  $y_{ijp}$  = the raw value for the well at row i and column j on plate p;  
 $\mu_p$  = an overall “average” for plate p;  
 $R'_{ip}$  = a possible systematic measurement row offset for row i on plate p;  
 $C'_{jp}$  = a possible systematic measurement column offset for column j on plate p; and  
25  $e_{ijp}$  = residual data without taking into account any non-additive interaction offset;

longitudinally (plate-wise) non-linearly smoothing each  $R'_{ip}$  and each  $C'_{jp}$ ;

30 substituting each un-smoothed  $R'_{ip}$  and  $C'_{jp}$  value with a corresponding smoothed  $R_{ip}$  and  $C_{jp}$  value and adjusting each residual value  $e_{ijp}$



21. The method of claim 17 comprising longitudinally (plate-wise) non-linearly smoothing each  $e'_{ijp}$  according to one of a running median smoother and a lowess procedure.

5

22. The method of claim 17 further comprising normalizing each  $r_{ijp}$  to result in a true relative value of the corresponding well (i, j, p) that can be compared to all other wells (i, j, p) on all plates p.

10

23. The method of claim 22 comprising normalizing each  $r_{ijp}$  by a standard deviation value derived from all the  $r_{ijp}$ 's on the plate p to result in a score for the well (i, j, p) that can be compared across plates p:

$$\text{score}_{ijp} = r_{ijp} / (\text{standard deviation value})_p .$$

15

24. The method of claim 23 comprising normalizing each  $r_{ijp}$  by a median absolute deviation from median value.

20

25. A computer having computer modules executing thereon for positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates p, each plate p having a plurality of wells organized into rows i and columns j, each well (i, j, p) having a raw value  $x_{ijp}$  associated therewith, the raw values  $x_{ijp}$  comprising the raw assay data, the modules comprising a first module deconstructing each raw value  $x_{ijp}$  of an associated well (i, j, p) into:

25

a plate effect value representing extraneous effects attributable to the plate p of the well (i, j, p);

30

a row effect value representing extraneous effects attributable to the row i on the plate p of the well (i, j, p);



a column effect value representing extraneous effects attributable to the column  $j$  on the plate  $p$  of the well  $(i, j, p)$ ;

a non-additive, interaction effect representing extraneous positional effects attributable to consistent positional effects beyond the plate, row, and column effects determined for the  $(i, j, p)$  well on plate  $p$ ; and

a residual data value that is left over once all the above extraneous effects are taken into account, the computer further comprising a second module employing the residual data value associated with each well  $(i, j, p)$  to represent the well  $(i, j, p)$  as compared with all other wells  $(i, j, p)$  on the plate  $p$ .

26. The computer of claim 25 wherein the raw assay data is generated from a high throughput screening assay to identify a biologically active agent in a collection of test agents, wherein a biologically active agent is identified by identifying a test agent that generates residual data value which statistically deviates from other residual data values generated.

27. The computer of claim 25 wherein the second module employs the residual data value associated with each well  $(i, j, p)$  to represent the well  $(i, j, p)$  as compared with all other wells  $(i, j, p)$  on all of the plates  $p$ .

28. The computer of claim 25 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well  $(i, j, p)$  into a plate effect value representing the overall median of all the raw values of the wells on plate  $p$ .

29. The computer of claim 28 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well  $(i, j, p)$  into a row effect value representing the median of all the raw values of the wells for row  $i$  on plate  $p$  after taking into consideration the plate effect value.

30. The computer of claim 28 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well (i, j, p) into a column effect value representing the median of all the raw values of the wells for column j on plate p after taking into consideration the plate effect value.

5

31. The computer of claim 25 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well (i, j, p) into a non-additive interactive effect value representing possible systematic measurement effect for the (i, j) well on plate p beyond that attributable to the plate, row and column effect values.

10

32. The computer of claim 25 further comprising an inputting module inputting the raw values  $x_{ijp}$  into a data structure in a memory of the computer, whereby the first module accesses the raw values  $x_{ijp}$  from the data structure.

15

33. A computer having computer modules executing thereon for positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates p, each plate p having a plurality of wells organized into rows i and columns j, each well (i, j, p) having a raw value  $x_{ijp}$  associated therewith, the raw values  $x_{ijp}$  comprising the raw assay data, the modules comprising:

20

a first module resistantly fitting the raw value  $x_{ijp}$  for each well (i, j) for each plate p to a row-column additive model:

25

$$y_{ijp} = \mu_p + R'_{ip} + C'_{jp} + e_{ijp} .$$

where:

30

$y_{ijp}$  = the raw value for the well at row i and column j on plate p;  
 $\mu_p$  = an overall "average" for plate p;

$R'_{ip}$  = a possible systematic measurement row offset for row  $i$  on plate  $p$ ;

$C'_{jp}$  = a possible systematic measurement column offset for column  $j$  on plate  $p$ ; and

5  $e_{ijp}$  = residual data without taking into account any non-additive interaction offset;

a second module longitudinally (plate-wise) non-linearly smoothing each  $R'_{ip}$  and each  $C'_{jp}$ ;

10 a third module substituting each un-smoothed  $R'_{ip}$  and  $C'_{jp}$  value with a corresponding smoothed  $R_{ip}$  and  $C_{jp}$  value and adjusting each residual value  $e_{ijp}$  to an adjusted  $e'_{ijp}$ :

$$y_{ijp} = \mu_p + R_{ip} + C_{jp} + e'_{ijp} ; \text{ and}$$

15 a fourth module longitudinally (plate-wise) non-linearly smoothing each  $e'_{ijp}$  to result in:

$$y_{ijp} = \mu_p + R_{ip} + C_{jp} + \text{smooth}_p(e'_{ijp}) + r_{ijp}$$

20 where each  $e'_{ijp}$  is deconstructed into  $\text{smooth}_p(e'_{ijp})$ , a possible systematic non-additive interaction offset for the  $(i, j)$  well on plate  $p$  and  $r_{ijp}$  residual data value left over after taking into account any non-additive interaction offset;

wherein each  $r_{ijp}$  represents a true relative value of the  
25 corresponding well  $(i, j, p)$  as compared to all other wells  $(i, j, p)$  on the plate  $p$ .

34. The computer of claim 33 wherein the raw assay data is generated from a high throughput screening assay to identify a biologically active agent in a collection of test agents, wherein a biologically active agent is identified  
30 by identifying a test agent that generates an  $r_{ijp}$  which statistically deviates from  $r_{ijp}$

generated.

35. The computer of claim 33 wherein the first module resistantly fits the raw value  $x_{ijp}$  for each plate  $p$  to a row-column additive model according to a two way resistant median polish procedure.

36. The computer of claim 33 wherein the second module longitudinally (plate-wise) non-linearly smoothes each  $R'_{ip}$  and each  $C'_{jp}$  according to one of a running median smoother and a lowess procedure.

37. The computer of claim 33 wherein the fourth module longitudinally (plate-wise) non-linearly smoothing each  $e'_{ijp}$  according to one of a running median smoother and a lowess procedure.

38. The computer of claim 33 further comprising a fifth module normalizing each  $r_{ijp}$  to result in a true relative value of the corresponding well  $(i, j, p)$  that can be compared to all other wells  $(i, j, p)$  on all plates  $p$ .

39. The computer of claim 38 wherein the fifth module normalizes each  $r_{ijp}$  by a standard deviation value derived from all the  $r_{ijp}$ 's on the plate  $p$  to result in a score for the well  $(i, j, p)$  that can be compared across plates  $p$ :

$$\text{score}_{ijp} = r_{ijp} / (\text{standard deviation value})_p .$$

40. The computer of claim 39 wherein the fifth module normalizes each  $r_{ijp}$  by a median absolute deviation from median value.

41. The computer of claim 33 further comprising an inputting module inputting the raw values  $x_{ijp}$  into a data structure in a memory of the

computer, whereby the first module accesses the raw values  $x_{ijp}$  from the data structure.

42. A computer-readable medium having computer-executable  
5 modules thereon for positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates  $p$ , each plate  $p$  having a plurality of wells organized into rows  $i$  and columns  $j$ , each well  $(i, j, p)$  having a raw value  $x_{ijp}$  associated therewith, the raw values  $x_{ijp}$  comprising the raw assay data, the modules comprising a first module for deconstructing each raw value  $x_{ijp}$   
10 of an associated well  $(i, j, p)$  into:  
a plate effect value representing extraneous effects attributable to the plate  $p$  of the well  $(i, j, p)$ ;  
a row effect value representing extraneous effects attributable to the row  $i$  on the plate  $p$  of the well  $(i, j, p)$ ;  
15 a column effect value representing extraneous effects attributable to the column  $j$  on the plate  $p$  of the well  $(i, j, p)$ ;  
a non-additive, interaction effect representing extraneous positional effects attributable to systematic positional effects beyond the plate, row, and column effects previously determined for the  $(i, j, p)$  well on plate  $p$ ; and  
20 a residual data value that is left over once all the above extraneous effects are taken into account,  
the computer further comprising a second module for employing the residual data value associated with each well  $(i, j, p)$  to represent the well  $(i, j, p)$  as compared with all other wells  $(i, j, p)$  on the plate  $p$ .

25

43. The computer-readable medium of claim 42 wherein the raw assay data is generated from a high throughput screening assay to identify a biologically active agent in a collection of test agents, wherein a biologically active agent is identified by identifying a test agent that generates a residual data value  
30 which statistically deviates from the other residual data value generated by the

assay.

44. The computer-readable medium of claim 42 wherein the second module employs the residual data value associated with each well (i, j, p) to represent the well (i, j, p) as compared with all other wells (i, j, p) on all of the plates p.

45. The computer-readable medium of claim 42 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well (i, j, p) into a plate effect value representing the overall median of all the raw values of the wells on plate p.

46. The computer-readable medium of claim 45 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well (i, j, p) into a row effect value representing the median of all the raw values of the wells for row i on plate p after taking into consideration the plate effect value.

47. The computer-readable medium of claim 45 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well (i, j, p) into a column effect value representing the median of all the raw values of the wells for column j on plate p after taking into consideration the plate effect value.

48. The computer-readable medium of claim 42 wherein the first module deconstructs each raw value  $x_{ijp}$  of an associated well (i, j, p) into a non-additive interactive effect value representing a possible systematic measurement effect for the (i, j) well on plate p with respect to (i, j) wells on nearby plates p.

49. The computer-readable medium of claim 42 further comprising an inputting module inputting the raw values  $x_{ijp}$  into a data structure in

a memory of a computer, whereby the first module accesses the raw values  $x_{ijp}$  from the data structure.

50. A computer-readable medium having computer-executable modules thereon for positionally correcting raw assay data from an assay comprising a plurality of longitudinally oriented plates p, each plate p having a plurality of wells organized into rows i and columns j, each well (i, j, p) having a raw value  $x_{ijp}$  associated therewith, the raw values  $x_{ijp}$  comprising the raw assay data, the modules comprising:

10 a first module for resistantly fitting the raw value  $x_{ijp}$  for each well (i, j) for each plate p to a row-column additive model:

$$y_{ijp} = \mu_p + R'_{ip} + C'_{jp} + e_{ijp} .$$

15 where:

$y_{ijp}$  = the raw value for the well at row i and column j on plate p;  
 $\mu_p$  = an overall “average” for plate p;  
 $R'_{ip}$  = a possible systematic measurement row offset for row i on plate p;

20  $C'_{jp}$  = a possible systematic measurement column offset for column j on plate p; and

$e_{ijp}$  = residual data without taking into account any non-additive interaction offset;

25 a second module for longitudinally (plate-wise) non-linearly smoothing each  $R'_{ip}$  and each  $C'_{jp}$ ;

a third module for substituting each un-smoothed  $R'_{ip}$  and  $C'_{jp}$  value with a corresponding smoothed  $R_{ip}$  and  $C_{jp}$  value and adjusting each residual value  $e_{ijp}$  to an adjusted  $e'_{ijp}$ ;

30

$$y_{ijp} = \mu_p + R_{ip} + C_{jp} + e'_{ijp} ; \text{ and}$$

a fourth module for longitudinally (plate-wise) non-linearly smoothing each  $e'_{ijp}$  to result in:

5

$$y_{ijp} = \mu_p + R_{ip} + C_{jp} + \text{smooth}_p(e'_{ijp}) + r_{ijp}$$

where each  $e'_{ijp}$  is deconstructed into  $\text{smooth}_p(e'_{ijp})$ , a possible systematic non-additive interaction measurement offset for the (i, j) well on plate p, and  $r_{ijp}$ ,  
10 residual data left over after taking into account any interaction offset;

wherein each  $r_{ijp}$  represents a true relative value of the corresponding well (i, j, p) as compared to all other wells (i, j, p) on the plate p.

51. The computer-readable medium of claim 50 wherein the raw  
15 assay data is generated from a high throughput screening assay to identify a biologically active agent in a collection of test agents, wherein a biologically active agent is identified by identifying a test agent that generates an  $r_{ijp}$  which statistically deviates from the  $r_{ijp}$  generated by the other agents in the assay.

20 52. The computer-readable medium of claim 50 wherein the first module resistantly fits the raw value  $x_{ijp}$  for each plate p to a row-column additive model according to a two way resistant median polish procedure.

53. The computer-readable medium of claim 50 wherein the  
25 second module longitudinally (plate-wise) non-linearly smoothes each  $R'_{ip}$  and each  $C'_{jp}$  according to one of a running median smoother and a lowess procedure.

54. The computer-readable medium of claim 50 wherein the  
30 fourth module longitudinally (plate-wise) non-linearly smoothing each  $e'_{ijp}$  according to one of a running median smoother and a lowess procedure.



55. The computer-readable medium of claim 50 further comprising a fifth module for normalizing each  $r_{ijp}$  to result in a true relative value of the corresponding well (i, j, p) that can be compared to all other wells (i, j, p) on all plates p.

56. The computer-readable medium of claim 55 wherein the fifth module normalizes each  $r_{ijp}$  by a standard deviation value derived from all the  $r_{ijp}$ 's on the plate p to result in a score for the well (i, j, p) that can be compared across plates p:

$$\text{score}_{ijp} = r_{ijp} / (\text{standard deviation value})_p .$$

57. The computer-readable medium of claim 56 wherein the fifth module normalizes each  $r_{ijp}$  by a median absolute deviation from median value.

58. The computer-readable medium of claim 50 further comprising an inputting module for inputting the raw values  $x_{ijp}$  into a data structure in a memory of a computer, whereby the first module accesses the raw values  $x_{ijp}$  from the data structure.

### Abstract of the Disclosure

5

An assay is performed such that a compendium of raw assay data is developed and is then positionally corrected. The assay comprises a plurality of longitudinally oriented plates  $p$ , each having a wells organized into rows  $i$  and columns  $j$ . Each well  $(i, j, p)$  has a raw value  $x_{ijp}$  associated therewith that is deconstructed into: a plate effect value representing extraneous effects attributable to the plate  $p$  of the well  $(i, j, p)$ ; a row effect value representing extraneous effects attributable to the row  $i$  on the plate  $p$  of the well  $(i, j, p)$ ; a column effect value representing extraneous effects attributable to the column  $j$  on the plate  $p$  of the well  $(i, j, p)$ ; a non-additive, interaction effect representing extraneous positional effects attributable to consistent positional effects beyond the plate, row, and column effects previously determined for the  $(i, j, p)$  well on plate  $p$ ; and a residual data value that is left over once all the above extraneous effects are taken into account. Thereafter, the residual data value associated with each well  $(i, j, p)$  is employed to represent the well  $(i, j, p)$  as compared with all other wells  $(i, j, p)$  on the plate  $p$ .

## APPENDIX

5

S-PLUS CODE FOR POSITIONALLY CORRECTING ALGORITHM  
version 1

10

```
"score.hts"<-  
15 function(obj.hts, format = if(!is.null(fmt <- attr(obj.hts, "format"))) fmt  
    else list(dim = c(8, 12), Rows = LETTERS[1:8], Columns = 2:11))  
  {  
    # obj.hts is an object of class "hts". This is a data.frame in which each  
    row  
20    # represents one well of results and must contain the following NAMED  
    columns:  
    # REQUIRED  
    # $Plate: Plate id (alphanumeric factor) IN THE ORDER THE PLATES WERE  
    RUN  
25    # $Row: Row id of the well (alphanumeric)  
    # $Col: Column id of the well (alphanumeric)  
    # $Type: Type of well contents:  
    # "D" for a sample compound or mixture  
    # "H" for high control (high raw measurements)  
30    # "L" for low control (low raw measurements)  
    # ... other alphanumeric codes for other possible controls  
    # $Value: RAW measured value (NOT %inhibition or excitation)  
    #  
    # OPTIONAL  
35    # $Run: Runset code  
    # $Date: The Date on which the sample was run  
    # $Samp.ID: The sample ID code (e.g., L-number)  
    #  
    # NOTE: The data (plates) must be given in the order they were run  
40    #  
    # More function arguments  
    # format: list with 3 components:  
    # $dim = c(number of rows,number of columns)in plate  
    # $Rows = the row id's of test samples
```

```

# $Columns = the column id's of test samples
# The defaults given are for 96 well plates (where controls are in
# columns 1 and 12)
#
5  # Make sure Value column is numeric. If not, stop with error message.
    if(!is.numeric(obj.hts$Value)) stop("Value column must be numeric.")
    nrow <- length(format$Rows)
    ncol <- length(format$Columns)
    nplate <- nrow * ncol
10   nm.obj <- names(obj.hts)
    # Make sure ordering of factors in data frame is maintained
    obj.hts$Plate <- ordered(obj.hts$Plate, levels =
unique(obj.hts$Plate))
    p.count <- table(obj.hts$Plate)
15   bad.plt <- p.count > prod(format$dim)
    if(any(bad.plt)) {
        cat("\n\t\t ***** Bad Plate Indexing *****\n\n
The following plate\n numbers appear more than once in the data:\n"
        )
20       bads <- p.count[bad.plt]
        cat("\tPLATE NUMB\t\tTotal Wells in Data with This Plate
Numb\n"
        )
        bads <- paste(" ", names(bads), round(bads, 0), sep =
25   "\t")
        cat(bads, "\n", sep = "\n")
        stop()
    }
    platelist <- as.vector(unique(obj.hts$Plate))
30   n.orig <- length(platelist)
    # Remove plates that are all controls, i.e. no sample wells("D") on
    them
    good.plates <- as.vector(unique(obj.hts$Plate[obj.hts$Type == "D"]))
    lnth <- length(good.plates)
35   if(lnth == n.orig)
        good.indx <- 1:n.orig
    else good.indx <- match(good.plates, platelist)
    plt.ind <- match(good.plates, obj.hts$Plate)
    #indices of good plates in plate column
40   if(lnth < length(p.count)) {
        obj.hts <- obj.hts[!is.na(match(obj.hts$Plate,
good.plates)),
        ]
        codes.new <- unique(codes(obj.hts$Plate))
45   obj.hts$Plate <- structure(match(codes(obj.hts$Plate),

```

```

        codes.new), levels =
levels(obj.hts$Plate)[codes.new],
        class = c("ordered", "factor"))
    }
5    pick.c <- c("Plate", "Row", "Col", "Value", "Samp.ID")
    if(is.na(match("Samp.ID", nm.obj)))
        samps <- obj.hts[obj.hts$Type == "D", pick.c[-5]]
    else samps <- obj.hts[obj.hts$Type == "D", pick.c]
    row <- match(samps$Row, format$Rows)
10   col <- match(samps$Col, format$Columns)
    if(any(is.na(row)))
        stop("Row codes for sample wells does not match format
specification."
    )
15   if(any(is.na(col)))
        stop("Column codes for sample wells does not match format
specification."
    )
    pl <- match(samps$Plate, good.plates)
20   # Fit an additive row/column fit for the samples on each plate
    y <- array(NA, c(nrow, ncol, lnth))
    samp.indx <- (pl - 1) * nplate + (col - 1) * nrow + row
    y[samp.indx] <- samps$Value
    fit.byplate <- apply(y, 3, function(x)
25   twoway(x)[-4])
    # Make sure row and column effects haven't been corrupted by a row # or
column with a majority of actives by smoothing
    unl <- unlist(fit.byplate, rec = F, use.n = F)
    rowfits <- unlist(unl[seq(2, by = 3, length = lnth)])
30   colfits <- unlist(unl[seq(3, by = 3, length = lnth)])
    grand <- unlist(unl[seq(1, by = 3, length = lnth)])
    na.smooth <- function(x, twice = T)
    {
    # If length>=10, smooth
35   if(sum(!is.na(x)) > 9) x[!is.na(x)] <- as.vector(smooth(x[!
        is.na(x)], tw = twice))
        x
    }
    rowfits <- matrix(rowfits, ncol = nrow, byrow = T)
40   colfits <- matrix(colfits, ncol = ncol, byrow = T)
    rowfits <- apply(rowfits, 2, na.smooth) # Smooth the row fits
    colfits <- apply(colfits, 2, na.smooth) # Smooth the column fits
    # Create array of fits; layers = plates
    fit.a <- array(apply(cbind(rowfits, colfits), 1, function(x, nr, n)
45   {

```

```

        outer(x[1:nr], x[(nr + 1):n], "+")
    }
    , nrow, nrow + ncol), dim = c(nrow, ncol, lnth)) + rep(grand, e =
        nplate)
5   resid.a <- y - fit.a #Residuals from smoothed row/column fits
#### Now smooth the residuals in each position over the Plate sequence
    if(lnth > 10) {
        resid.sm <- aperm(apply(resid.a, 1:2, na.smooth, twice = F),
10   c(
            2, 3, 1))
        class(resid.sm) <- NULL
        fit.a <- fit.a + resid.sm # final overall fits
        resid.a <- resid.a - resid.sm # final residuals
    }
15   names(resid.a) <- NULL
    ##### Obtain the spread of the finals residuals on each plate as the mads
    mads <- apply(resid.a, 3, function(x, np)
    {
        mad(x, na = T) * sqrt(np/sum(!is.na(x))) #df correction for missing
20   values
    }
    , nplate)
    # convert 0 mads to NA's to indicate that reasonable scores can't be
    computed
25   mads[mads == 0] <- NA
    # Compute means of low and high controls and activities
    act <- tapply(1:(dim(obj.hts)[1]), obj.hts$Plate, function(x, z, type)
    {
        z <- z[x]
30   type <- type[x]
        lo <- mean(z[type == "L"], na.rm = T)
        hi <- mean(z[type == "H"], na.rm = T)
        diff <- hi - lo
        if(diff > 0)
35   act <- (100 * (z - lo))/diff
        else act <- rep(NA, length(z))
        list(lo, hi, act)
    }
    , obj.hts$Value, obj.hts$Type)
40   unl <- unlist(act, rec = F, use.n = F)
    lows <- unlist(unl[seq(1, by = 3, length = n.orig)])[good.indx]
    highs <- unlist(unl[seq(2, by = 3, length = n.orig)])[good.indx]
    act <- unlist(unl[seq(3, by = 3, length = n.orig)])
    # Determine whether high or low controls are potent
45   if(any(!is.na(lows)) && any(!is.na(highs))) {

```

```

        mlo <- median(lows, na.rm = T)
        mhi <- median(highs, na.rm = T)
        mgr <- median(grand, na.rm = T)
        if(abs(mlo - mgr) > abs(mhi - mgr))
5           potent <- "low"
        else potent <- "h"
    }
    else potent <- NA
    scores <- (resid.a/rep(mads, e = nplate)) # B-scores for all sample wells
10    bsc <- rep(NA, e = dim(obj.hts)[1])
    bsc[obj.hts$Type == "D"] <- scores[samp.indx]
    # Housekeeping to track runs, number of plates per run, etc.
    if(is.na(match("Run", nm.obj)) |
length(unique(obj.hts$Run[plt.ind])) ==
15         1) {
        Run <- Inth
        names(Run) <- "All"
        runset <- rep(1, Inth)
    }
    else {
20         runset <- obj.hts$Run[plt.ind]
        Run <- table(runset)
    }
    if(!is.na(match("Date", nm.obj)))
25         date <- as.character(obj.hts$Date[plt.ind])
    else date <- rep(NA, Inth)
    if(is.na(match("Samp.ID", nm.obj)))
        Samp.ID <- NA
    else {
30         Samp.ID <- array(NA, c(nrow, ncol, Inth))
        Samp.ID[samp.indx] <- sampls$Samp.ID
    }
    the.call <- sys.call()
    out <- structure(list(Call = the.call, Format = format, N.orig = n.orig,
35         Potent = potent, Run = Run, Plate.stats = data.frame(Plate =
        good.plates, Date = date, Runset = runset, Center = grand,
        Scale = mads, Low.cntl = lows, Hi.cntl = highs, row.names = rep(
        NA, Inth), dup.row.names = T), Effects = list(Roweff = rowfits,
        Coleff = colfits), Results = list(Fitted = fit.a, Resid =
40         resid.a, Samp.ID = Samp.ID), Activity = act, Bscore = bsc),
        class = "htsfit")
    if(all(is.na(out$Plate.stats$Low.cntl)) ||
all(is.na(out$Plate.stats$
        Hi.cntl)))
45         class(out) <- c("bscr.only", class(out))

```

invisible(out)

}

\*\*\*\*\*

[illegible]



5 S-PLUS CODE FOR POSITIONALLY CORRECTING ALGORITHM  
version 2

```

10 "score.hts"<-
function(obj.hts, format = if(!is.null(fmt <- attr(obj.hts, "format"))) fmt
      else list(dim = c(8, 12), Rows = LETTERS[1:8], Columns = 2:11),
plate,
15   plate.span = if(prod(format$dim) == 96) list(effects = 15, resids =
      11,
          mads = 15) else list(effects = 11, resids = 11, mads = 11))
{
# obj.hts is an object of class "hts". This is a data.frame in which each
20 row
# represents one well of results and must contain the following NAMED
columns:
# REQUIRED
# $Plate: Plate id (alphanumeric factor) IN THE ORDER THE PLATES WERE
25 RUN
# $Row: Row id of the well (alphanumeric)
# $Col: Column id of the well (alphanumeric)
# $Type: Type of well contents:
# "D" for a sample compound or mixture
30 # "H" for high control (high raw measurements)
# "L" for low control (low raw measurements)
# ... other alphanumeric codes for other possible controls
# $Value: RAW measured value (NOT %inhibition or excitation)
#
35 # OPTIONAL
# $Run: Runset code
# $Date: The Date on which the sample was run
# $Samp.ID: The sample ID code (e.g., L-number)
#
40 # NOTE: The data (plates) must be given in the order they were run
#
# More function arguments
# format: list with 3 components:
# $dim = c(number of rows,number of columns)in plate
45 # $Rows = the row id's of test samples

```

```

# $Columns = the column id's of test samples
# The defaults given are for 96 well plates (where controls are in
# columns 1 and 12)
#
5  # Make sure Value column is numeric. If not, stop with error message.
    if(!is.numeric(obj.hts$Value)) stop("Value column must be numeric.")
    nrow <- length(format$Rows)
    ncol <- length(format$Columns)
    nplate <- nrow * ncol
10  nm.obj <- names(obj.hts)
    # Make sure ordering of factors in data frame is maintained
    obj.hts$Plate <- ordered(as.character(obj.hts$Plate), levels =
unique(
    obj.hts$Plate))
15  p.count <- table(obj.hts$Plate)
    bad.plt <- p.count > prod(format$dim)
    if(any(bad.plt)) {
        bads <- p.count[bad.plt]
        bads <- paste(" ", names(bads), " ", round(bads, 0),
20  sep
            = "\t", collapse = "\n")
        stop(paste("\n\t\t ***** Bad Plate Indexing
*****\n \n The following plate numbers appear more than once in the
data:\n\n \tPLATE NUMBER\tTotal Wells in Data with This Plate Number\n",
25  bads, sep = ""))
    }
    platelist <- as.vector(unique(obj.hts$Plate))
    n.orig <- length(platelist)
    # Remove plates that are all controls, i.e. no sample wells("D") on
30  them
        good.plates <- as.vector(unique(obj.hts$Plate[obj.hts$Type == "D"]))
        lnth <- length(good.plates)
        if(lnth == n.orig)
            good.indx <- 1:n.orig
35  else good.indx <- match(good.plates, platelist)
        indexofsamps <- obj.hts$Type == "D"
        if(lnth < length(p.count)) {
            obj.hts <- obj.hts[!is.na(match(obj.hts$Plate,
40  good.plates)),
                ]
            codes.new <- unique(codes(obj.hts$Plate))
            obj.hts$Plate <- structure(match(codes(obj.hts$Plate),
                codes.new), levels =
levels(obj.hts$Plate)[codes.new],
45  class = c("ordered", "factor"))

```

```

    }
    plt.ind <- match(good.plates, obj.hts$Plate)
    #indices of good plates in plate column
    pick.c <- c("Plate", "Row", "Col", "Value", "Samp.ID")
5    if(is.na(match("Samp.ID", nm.obj)))
        samps <- obj.hts[obj.hts$Type == "D", pick.c[-5]]
    else samps <- obj.hts[obj.hts$Type == "D", pick.c]
    row <- match(samps$Row, format$Rows)
    col <- match(samps$Col, format$Columns)
10    if(any(is.na(row)))
        stop("Row codes for sample wells does not match format
specification."
    )
    if(any(is.na(col)))
15    stop("Column codes for sample wells does not match format
specification."
    )
    pl <- match(samps$Plate, good.plates)
    # Fit an additive row/column fit for the samples on each plate
20    y <- array(NA, c(nrow, ncol, lnth))
    samp.indx <- (pl - 1) * nplate + (col - 1) * nrow + row
    y[samp.indx] <- samps$Value
    fit.byplate <- apply(y, 3, function(x)
25    twoway(x, trim = 0.15)[-4])
    # Make sure row and column effects haven't been corrupted by a row
# or column with a majority of actives by smoothing
    unl <- unlist(fit.byplate, rec = F, use.n = F)
    rowfits <- unlist(unl[seq(2, by = 3, length = lnth)])
    colfits <- unlist(unl[seq(3, by = 3, length = lnth)])
30    grand <- unlist(unl[seq(1, by = 3, length = lnth)])
    na.smooth <- function(x, span, delta = 2, method = "lowess")
    {
    # The method= argument gives added resistance
    # If length>span, smooth nonmissings
35    ok <- !is.na(x)
    nok <- sum(ok)
    n <- length(x)
    if(nok > span) {
        if(method == "tukey")
40    x[ok] <- as.vector(smooth(x[ok], twice = F))
        span <- min(nok/2, span)
        delta <- min(delta, 0.01 * nok)
        x[ok] <- lowess((1:n)[ok], x[ok], f = span/nok,
delta
45    = delta)$y

```

```

        }
        x
    }
    rowfits <- matrix(rowfits, ncol = nrow, byrow = T)
5   colfits <- matrix(colfits, ncol = ncol, byrow = T)
    # Smooth the row fits
    spn <- plate.span$effects
    rowfits <- apply(rowfits, 2, na.smooth, span = spn)
    # Smooth the column fits
10   colfits <- apply(colfits, 2, na.smooth, span = spn)
    # Create array of fits; layers = plates
    fit.a <- array(apply(cbind(rowfits, colfits), 1, function(x, nr, n)
    {
        outer(x[1:nr], x[(nr + 1):n], "+")
15   }
        , nrow, nrow + ncol), dim = c(nrow, ncol, lnth)) + rep(grand, e =
            nplate)
    resid.a <- y - fit.a    #Residuals from smoothed row/column fits
##### Now smooth the residuals in each position over the Plate sequence
20   if(lnth > 10) {
        resid.sm <- aperm(apply(resid.a, 1:2, na.smooth, span =
            plate.span$resids, method = "tukey"), c(2, 3, 1))
        class(resid.sm) <- NULL
        fit.a <- fit.a + resid.sm    # final overall fits
25   resid.a <- resid.a - resid.sm    # final residuals
    }
    names(resid.a) <- NULL
##### Obtain the spread of the finals residuals on each plate
as the mads
30   mads <- apply(resid.a, 3, mad, na.rm = T)
    # convert 0 mads to NA's to indicate that reasonable scores can't be
    computed
    mads[mads == 0] <- NA
    mads <- exp(na.smooth(log(mads), span = plate.span$mads))
35   # Compute means of low and high controls and activities
    act <- tapply(1:(dim(obj.hts)[1]), obj.hts$Plate, function(x, z,
type)
    {
        z <- z[x]
40   type <- type[x]
        lo <- mean(z[type == "L"], na.rm = T)
        hi <- mean(z[type == "H"], na.rm = T)
        diff <- hi - lo
        z <- z[type == "D"]
45   if(diff > 0)

```

```

        act <- (100 * (z - lo))/diff
        else act <- rep(NA, length(z))
        list(lo, hi, act)
    }
5    , obj.hts$Value, obj.hts$Type)
    unl <- unlist(act, rec = F, use.n = F)
    lows <- unlist(unl[seq(1, by = 3, length = lnth)])
    highs <- unlist(unl[seq(2, by = 3, length = lnth)])
    act <- unlist(unl[seq(3, by = 3, length = lnth)])
10   # Determine whether high or low controls are potent
    if(any(!is.na(lows)) && any(!is.na(highs))) {
        mlo <- median(lows, na.rm = T)
        mhi <- median(highs, na.rm = T)
        mgr <- median(grand, na.rm = T)
15   if(abs(mlo - mgr) > abs(mhi - mgr))
        potent <- "low"
        else potent <- "h"
    }
    else potent <- NA
20   scores <- (resid.a/rep(mads, e = nplate))
    # B-scores for all sample wells
    bsc <- rep(NA, e = sum(p.count))
    Act <- bsc
    bsc[indexofsamps] <- scores[samp.indx]
25   Act[indexofsamps] <- act
    # Housekeeping to track runs, number of plates per run, etc.
    if(is.na(match("Run", nm.obj)) |
length(unique(obj.hts$Run[plt.ind])) ==
        1) {
30   Run <- lnth
        names(Run) <- "All"
        runset <- rep(1, lnth)
    }
    else {
35   runset <- as.character(obj.hts$Run[plt.ind])
        ## use unique() to assure correct ordering in table()
        Run <- table(runset)[unique(runset)]
    }
    if(!is.na(match("Date", nm.obj)))
40   date <- as.character(obj.hts$Date[plt.ind])
    else date <- rep(NA, lnth)
    if(is.na(match("Samp.ID", nm.obj)))
        Samp.ID <- NA
    else {
45   Samp.ID <- array(NA, c(nrow, ncol, lnth))

```

```
        Samp.ID[samp.indx] <- as.character(samps$Samp.ID)
    }
    the.call <- sys.call()
    out <- structure(list(Call = the.call, Format = format, N.orig =
5    n.orig,
        Potent = potent, Run = Run, Plate.stats = data.frame(Plate =
        good.plates, Date = date, Runset = runset, Center = grand,
        Scale = mads, Low.cntl = lows, Hi.cntl = highs, row.names =
10    rep(
        NA, Inth), dup.row.names = T), Effects = list(Roweff =
        rowfits,
        Coleff = colfits), Results = list(Fitted = fit.a, Resid =
        resid.a, Samp.ID = Samp.ID), Activity = Act, Bscore = bsc),
15    class = "htsfit")
        if(all(is.na(out$Plate.stats$Low.cntl)) ||
        all(is.na(out$Plate.stats$
        Hi.cntl)))
        class(out) <- c("bscr.only", class(out))
20    invisible(out)
    }

25
```

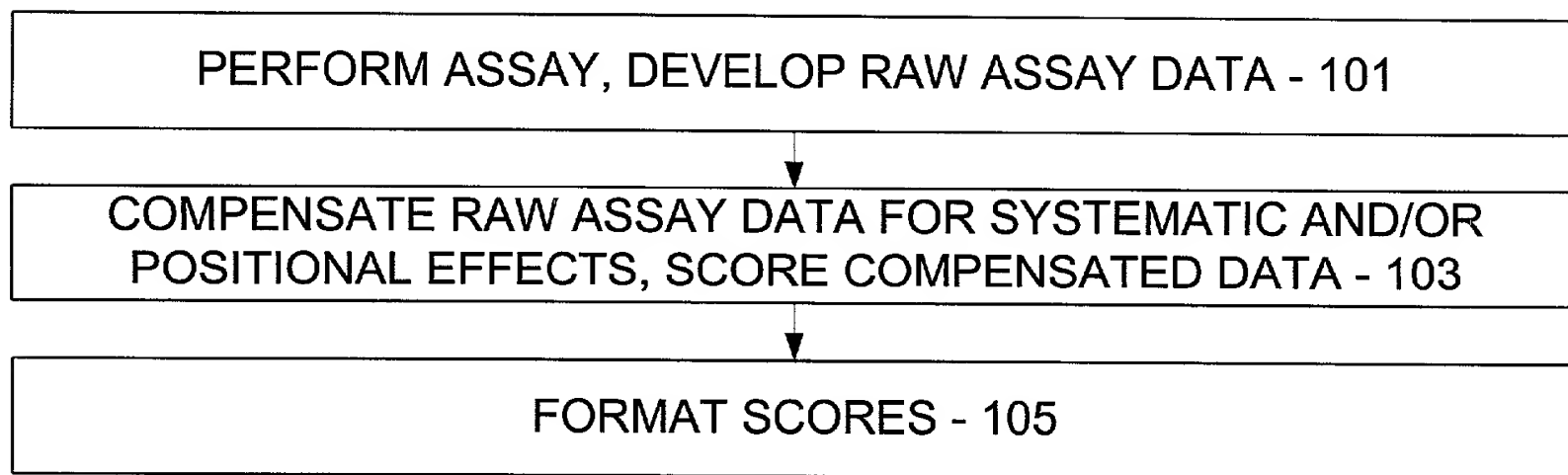


FIG. 1

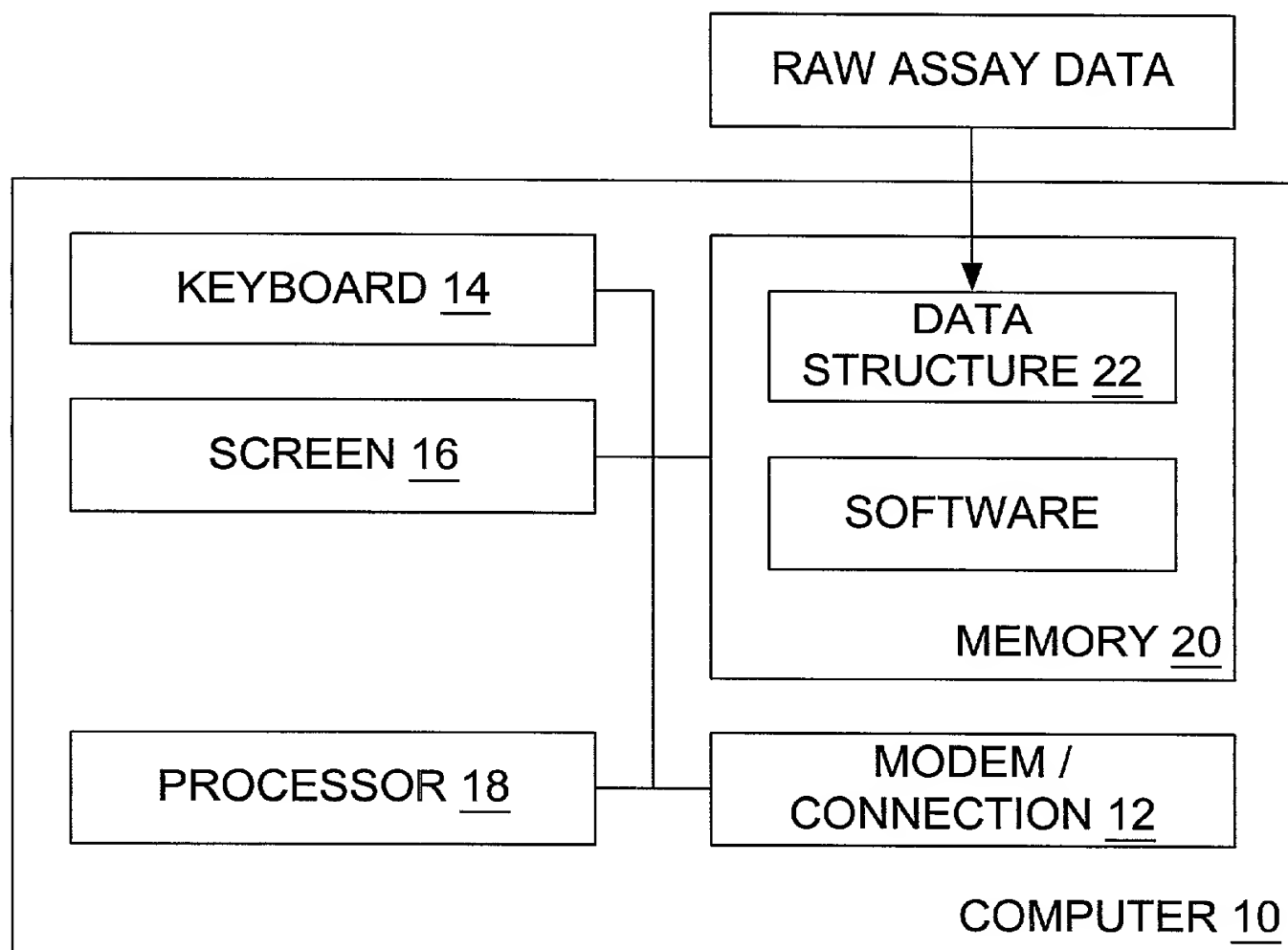


FIG. 2

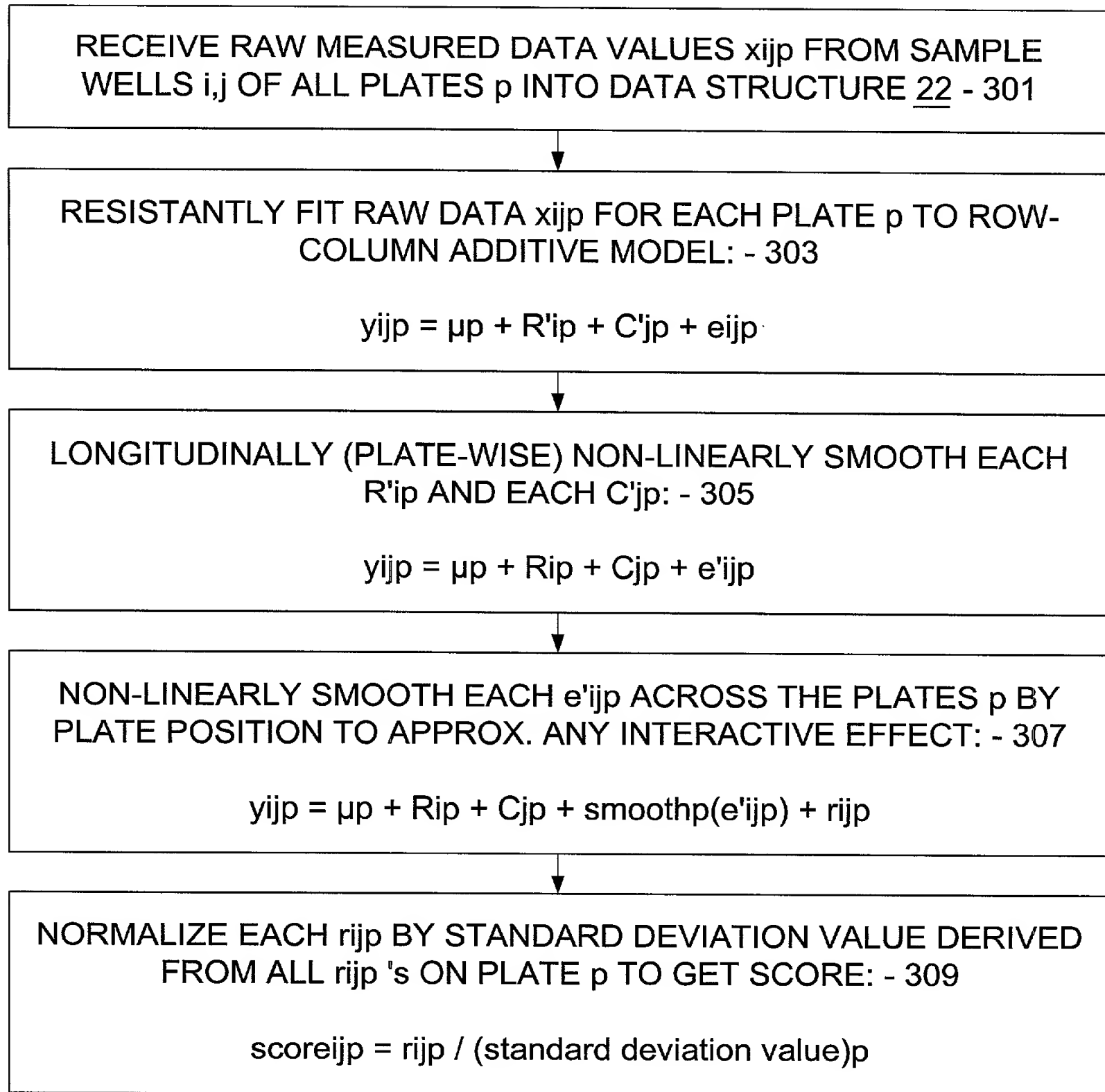


FIG. 3



## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

## In Re Application of:

Bert Gunter

EXPRESS MAIL LABEL NO:  
EL531431154USGroup Art Unit: Not Yet  
AssignedFor: METHOD AND APPARATUS FOR  
POSITIONALLY CORRECTING DATA  
IN A THREE DIMENSIONAL ARRAY

Examiner: Not Yet Assigned

## DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name; and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a

☒ Utility Patent ☐ Design Patent

is sought on the invention, whose title appears above, the specification of which:

☒ is attached hereto.  
☐ was filed on \_\_\_\_\_ as Serial No. \_\_\_\_\_.  
☐ said application having been amended on \_\_\_\_\_.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to be material to the patentability of this application in accordance with 37 CFR § 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. § 119(a-d) of any foreign

**application(s)** for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of any application on which priority is claimed:

Priority Claimed (If X'd)	Country	Serial Number	Date Filed
<input type="checkbox"/>	_____	_____	_____
<input type="checkbox"/>	_____	_____	_____
<input type="checkbox"/>	_____	_____	_____
<input type="checkbox"/>	_____	_____	_____

I hereby claim the benefit under 35 U.S.C. § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of 35 U.S.C. § 112, I acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to be material to patentability as defined in 37 CFR § 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

Serial Number	Date Filed	Patented/Pending/Abandoned
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

I hereby claim the benefit under 35 U.S.C. § 119(e) of any United States provisional application(s) listed below:

Serial Number	Date Filed
<u>60/217,772</u>	<u>July 12, 2000</u>

I hereby appoint the following persons as attorney(s) and/or agent(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Elliott Korsen

Registration No. 32,705

of **MERCK & Co., Inc.**, P.O. Box 2000, RY60-30, Rahway, NJ 07065 and

Mark DeLuca

Registration No. 33,229

Steven H. Meyer

Registration No. 37,189

of **WOODCOCK WASHBURN KURTZ MACKIEWICZ & NORRIS LLP**, One Liberty Place - 46<sup>th</sup> Floor, Philadelphia, Pennsylvania 19103

Please address all telephone calls and correspondence to:

Steven H. Meyer

**WOODCOCK WASHBURN KURTZ  
MACKIEWICZ & NORRIS LLP**

One Liberty Place - 46<sup>th</sup> Floor

Philadelphia PA 19103

Telephone: (215) 568-3100

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

<b>Name:</b> Bert Gunter	
<b>Mailing Address:</b>	<b>Signature</b>
<b>City/State of Actual Residence:</b>	<b>Date of Signature:</b> _____
	<b>Citizenship:</b> _____

<b>Name:</b>	
<b>Mailing Address:</b>	<b>Signature</b>
<b>City/State of Actual Residence:</b>	<b>Date of Signature:</b> _____
	<b>Citizenship:</b> _____

<b>Name:</b>	
<b>Mailing Address:</b>	<b>Signature</b>
<b>City/State of Actual Residence:</b>	<b>Date of Signature:</b> _____
	<b>Citizenship:</b> _____